PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

Commissioner
US Department of Commerce
United States Patent and Trademark Office, PCT
2011 South Clark Place Room CP2/5C24
Arlington, VA 22202
ETATS-UNIS D'AMERIQUE

in its capacity as elected Office

Date of mailing: 15 March 2001 (15.03.01)	
International application No.: PCT/AU00/01023	Applicant's or agent's file reference: 127064
International filing date: 30 August 2000 (30.08.00)	Priority date: 06 September 1999 (06.09.99)
Applicant: COX, Michael, S. et al	

1. The designated Office is hereby notified of its election made:

[X] in the demand filed with the International preliminary Examining Authority on:

15 December 2000 (15.12.00)

[ ] in a notice effecting later election filed with the International Bureau on:

2. The election [X] was

[ ] was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No.: (41-22) 740.14.35	Authorized officer: J. Zahra Telephone No.: (41-22) 338.83.38

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 March 2001 (15.03.2001)

PCT

(10) International Publication Number
WO 01/19019 A1

(51) International Patent Classification[7]: H04L 12/00

(21) International Application Number: PCT/AU00/01023

(22) International Filing Date: 30 August 2000 (30.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
47400/99 6 September 1999 (06.09.1999) AU

(71) Applicant (*for all designated States except US*): ALCATEL [FR/FR]; 54, rue la Boétie, F-75009 Paris (FR).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): COX, Michael, S. [AU/AU]; 101 7th Avenue, Jannali, NSW 2226 (AU). VUCIC, Mickey [AU/AU]; 25 Saiala Road, Killara, NSW 2071 (AU). BANH, Bui, Anh, Jonathan [AU/AU]; 34 Meryla Street, Burwood, NSW 2134 (AU).

(74) Agent: CONRICK, Patrick, Michael; Intellectual Property Dept., Alcatel Australia Limited, 280 Botany Road, Alexandria, NSW 2015 (AU).

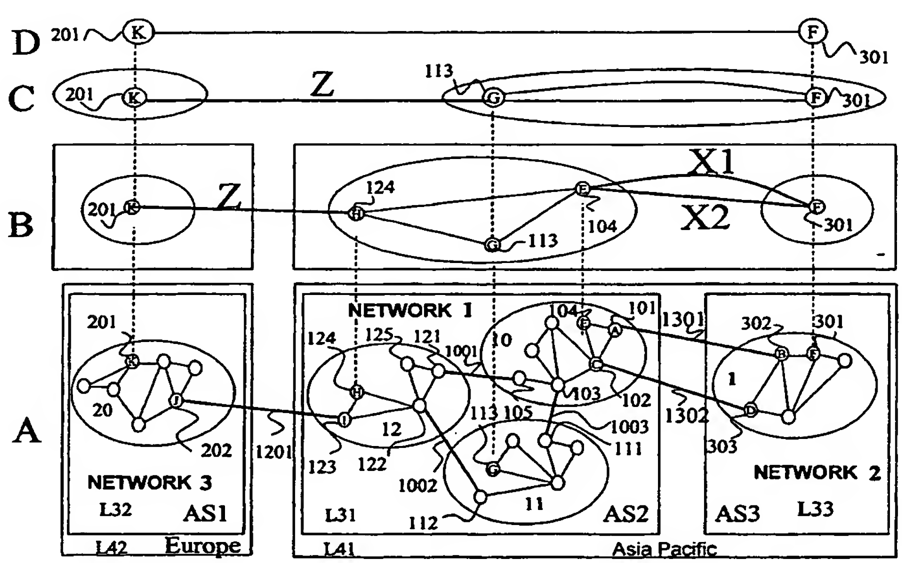
(81) Designated States (*national*): AU, CN, US.

(84) Designated States (*regional*): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published:
— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: RECURSIVE TRAFFIC DISTRIBUTION IP/DATA NETWORK MODEL



(57) Abstract: To avoid the need for all the nodes of a network to know the load status of all other nodes and links in the network, the nodes (101....303) are organized in node groups (10, 11, 12, 20, 30). Each group has a master node (104, 124, 113, 201, 301) which is incorporated in a logical higher order group. By recursively grouping the units of each preceding group, a logical hierarchical structure is formed in which the number of units at each level (A, B, C, D), decreases. Each node informs the other nodes of its load status, and each higher level unit also exchanges information with the other units. The recursive, hierarchical structure greatly reduces the amount of load status information exchanged across the network.

WO 01/19019 A1

RECURSIVE TRAFFIC DISTRIBUTION IP/DATA NETWORK MODEL

Technical field

This invention relates to a method and arrangement for transferring network node status information between nodes.

Background Art

Least cost routing examines the network topography to determine the shortest path between the source of a message and its destination. This method does not take account of the load status of the individual links of the chosen path so one or more of the links may become overloaded, preventing or disrupting the delivery of the message to the destination.

An alternative proposal is to take into account the load status of the links when determining the chosen path. Thus the system may determine all the shortest paths and make the path selection on the basis of the path with the links carrying the least traffic. To implement this, it is necessary for all the nodes to exchange load status information. As the number of nodes and links in a network grow, the implementation of this technique requires the exchange of a large amount of load status traffic, as each node broadcasts the load status of its associated links to the other nodes.

In our co-pending application number 44470/99 (127045 SY), we disclose a network in which the message is spread over all the practical paths between the source and destination. An advantageous embodiment of that invention involves the elimination of paths having heavily loaded links. Again, the implementation of this technique requires the exchange of load status information between nodes, generating a large volume of traffic.

Disclosure of the Invention

This specification discloses a network arrangement for a plurality of nodes each node being connected to one or more other nodes by corresponding node links,

- the network being arranged into a recursive hierarchy of units having two or more levels,
- the nodes being the units of the first level of the hierarchy,

- the units of higher levels of the hierarchy being formed by groupings of the units of the previous level,
- wherein the units of a level exchange a corresponding load status information.

This invention can be used in conjunction with the maximal flow techniques described in 44470/99 to determine suitable paths with available capacity.

Brief Description of the Drawings

Figure 1 is a schematic representation of a network in which the nodes are arranged in a recursive hierarchy, in accordance with an embodiment of the invention.

Figure 2 represents a load status monitor.

Figure 3 illustrates the message structure for exchanging information at different levels.

Best Mode of Carrying out the Invention

Figure 1 shows a network of nodes interconnected by links. According to the embodiment shown in Figure 1, the nodes are linked in a logical hierarchy.

The nodes exemplified by the circles, some of which are numbered 101 ... 202 ... 303 ..., are shown interconnected by node links, represented by the lines drawn between the nodes.

The nodes are formed into groups 10, 11, 12, 20, 30. The groups are interconnected by group links, for example 1001 between node 105 of group 10 and node 121 of group 12.

The group links are shown in the following Table 1.

TABLE 1 : GROUP LINKS

Node/Group		Node/Group		Link
105	10	121	12	1001
112	11	122	12	1002
103	10	111	11	1003
123	12	202	20	1201
101	10	302	30	1301
102	10	303	30	1302

Preferably the group links have a larger traffic capacity than node links. Group links may equate to regional trunks within a particular carrier's network, or to links between different carriers, different countries or different global regions, for example.

As shown in Figure 1, the levels of the hierarchy start with the nodes. The units of the second level are the groups of nodes. The units L31, L32, L33, of the third level are one or more groups, and the units of the fourth level, L41, L42 are formed by aggregating units of the third level. Thus L41 is the aggregation of L31 and L33, while L42 encompasses L32.

Preferably the nodes are grouped on the basis of communication path topography so that there are relatively few links in the shortest path between any two nodes in a group.

The groups themselves, 10, 11, 12, 20, 30 are linked by designated links 1001, 1002, 1003 1201, 1301, 1302 connecting designated nodes in the corresponding groups.

Within each group, a master node is assigned, 104, 113, 124, 201, 301. Because each node in a group knows the load status of all the links in that group, the master node has the information to enable it to compile available capacity reports adapted to meet the requirements of the higher levels.

The physical node interconnections are illustrated at A in Figure 1 and B,C and D illustrate the conceptual logical links for the second, third and fourth levels of hierarchy.

At level A, the nodes of each group communicate load status information to each of the other nodes of the corresponding group.

Thus the master nodes 104, 113, 124, of the network 1 shown at B are aggregated under the supervision of a single master node 113 which is assigned to the next higher level. Because, in the example shown, networks 2 and 3 have only one group, the same master 201, 301 is used all levels.

At level C, the master nodes 113, 201, 301 exchange information as to the overall load status of their associated networks 1, 2, and 3.

The nodes at level C are then 113, 201, 301. By then aggregating 113 and 301, one of these two can be designated to manage the level D information exchange for both network 1 and network 2. Thus at level D, 201 and 301 can exchange information on the available capacity for the regions covered by L42 and L41.

At level C, information on the capacity of networks 1, 2 and 3 is exchanged between the networks . At level B, information on the capacity of groups 10, 11, 12, 20, 30 is exchanged between the groups.

At level A, the nodes interchange information on traffic capacity at the node link level, within the groups.

Figure 2 shows an arrangement for monitoring the available capacity of the links connected to a node.

For each link connected to a node there is a buffer 51, 52, 53, e.g. in the form of FIFO.

The traffic level monitor 50 checks the level of the contents of the buffers to measure the available capacity on the basis of the speed of the link associated with the buffer. The result of the monitoring is then reported to the other nodes in the same group.

In a simplified measuring system the monitor may report whether or not a link has spare capacity, e.g. by checking whether a buffer's content is above or below a predetermined threshold.

The load status information exchange is carried out on the following basis.

The nodes within a group each notify the other nodes within that group of the load status of the links connected to the notifying node.

At the group level, each group notifies the other groups of the load status of the links connected to the notifying group and a summary of the load status of internal paths within the group available for interconnecting the group links connected to the notifying group.

For example, Group 12 is connected to Group 20 via link 1201, to Group 10 via link 1001, and to Group 11 via link 1002.

Preferably, the designated as a master node manages the interchange of information between the groups.

Table 2 shows the master nodes for each group.

TABLE 2

GROUP	MASTER NODE
10	104
11 -	113
12	124
20	201
30	301

As can be seen in Figure 1, the master nodes take part in the higher level exchanges but their number is progressively reduced by the recursive grouping.

Thus, in the embodiment shown in Figure 1, while there are 5 master nodes shown at level B, there are only 3 at level C and 2 at level D.

Preferably, the grouping is carried out on the basis of proximity in the sense of the number of links in the path. Of course this is not a strict rule at the node level because the nodes at either end of a group link are joined by a single link, while there may be more than 2 links between nodes within a group. Other factors which influence grouping are geographical proximity and network ownership, as well as the traffic flows.

For example, the nodes of network 2 may be geographically close to node of network 1, but network 1 may be owned by a different carrier from network 2.

At level D, the nodes 201 and 301 exchange information on the available capacity between network 3 and network 2 and the transit capacity of the respective networks. This information would, for example, be based on the load status of links 1201, 1301, 1302, and the capacity across network 1 between link 1201 and the links 1301, 1302. The information need only identify the maximum available capacity at the time, which varies in accordance with the load on the various network elements.

For the sake of clarity the information will be given the following names:

Level D = Regional;

Level C = Network;

Level B = Group;

Level A = Node.

Regional information may be, for example, the maximum available capacity between the "electrically" remotest groups. The term "electrically" refers to the number of links and may include cable, optical and radio links.

Network information may be, for example, the capacity between the various networks, including the trans-network capacity between the network links 1201, 1301, 1302.

Group information could be typified by the capacity between groups, including the trans-group capacity between the group links.

Node information is the information broadcast by a node to the other nodes within its group as the load status of the node and its associated links.

Group information can be deduced from node information. Each node in a group knows the load status of all the nodes in that group. Thus the master none 124 in group 12 knows the status of group links 1001 from node 121, group link 1002 from node 122, and group/network link 1201 from node 123, as well as the status of all the internal nodes and links within group 12. Node 124 can therefore calculate the available capacity across the group 12 between any pair of the links 1201, 1001, 1002. Preferably the master node 124 would use the "all practical paths" algorithm of our Australian Patent application 44470/99 (Docket No. 127045 SY) to calculate the trans-group capacity. This group information is interchanged between the group master nodes 201, 124, 113, 104, 301 at level B.

The units of the level B group domain are again grouped together, in this embodiment, into 3 network groups. The network groups include two one member groups 201 and 301, and one three member group 124, 113, 104. The network master of each one member group is the member of the group, while 113 is designated as the master of the three member network group.

The three network masters from level B interchange network information at the level C network domain. The information relates to the network links connecting the respective networks, and the trans-network information relating to the capacity between the pairs of network links.

At the regional domain, level D, the network masters 201, 113, 301 have been formed into two groups, resulting in two regional masters 201, 301, which exchange information on the available capacity between the two regions.

The regional master nodes 201, 301, convey the regional link capacity information to the other regional nodes. In the present embodiment 301 conveys the information to 113. 201 is the only regional node in the other regional grouping.

The regional nodes 201, 113 and 301 are all network master nodes and they convey the inter-regional and inter-network capacity information to the network level nodes. In our embodiments, 113 conveys this information to the nodes 104, 124.

Each of the network level nodes 201, 124, 113, 104, 301 is a group master and relays the higher level information to each of the nodes in its group.

The grouping of the units at each level means that the information exchanged at each level becomes more generalised.

This means that a node has detailed capacity information about the other nodes in its group. Capacity information about other groups in its network, capacity information about the other networks in its region, and information about the inter-regional capacity.

In a preferred embodiment the group master handles the interchange of node link capacity information. Each node, instead of broadcasting its load status to all the other nodes in the group, sends the information only to the group master, which collates the information from each node and relays the information to the other nodes. The message from the group master preferably incorporates the higher level load status information, so that each node has an overall picture of the entire system.

Thus the group master may broadcast a message including the information shown in Figure 3. The first segment RL includes the load status at the regional link level D. A second portion of the payload includes a number of segments of

information on the inter-network load status NL. A third portion includes segments GL on the inter-group load status, and the fourth portion includes segments NL on the load status of the nodes within the group. Alternatively, this information can be flooded to other part of the network using other means, such as a broadcast or multicast mechanism.

The claims defining the invention are as follows:

1. A network arrangement for a plurality of nodes each node being connected to one or more other nodes by corresponding node links,

- the network being arranged into a recursive hierarchy of units having two or more levels,
- the nodes being the units of the first level of the hierarchy,
- the units of higher levels of the hierarchy being formed by groupings of the units of the previous level,
- wherein the units of a level exchange a corresponding load status information.

2. An arrangement as claimed in claim 1 wherein, within, each group of units, a master entity is designated, the master entity conveying inter-unit load status information relating to the units of that level to the next higher level.

3. An arrangement as claimed claim 1 or claim 2 wherein, in the first level, a selected node in each group is designated as the master node for the corresponding group,

- the master node managing the transfer of node load status information within its corresponding group.

4. An arrangement as claimed in claim 1 or claim 2, or claim 3 wherein the load status information includes information on the available traffic capacity between the ports of each unit.

5. An arrangement as claimed in any one of claims 1 to 4 wherein each node includes node load status monitoring means to monitor the load status of the links connected to the node.

6. An arrangement as claimed in any one of claims 1 to 5 wherein at least one node of each second level group is connected to a node of at least one other second level group via a corresponding group link whereby group load status information can be interchange.

7. An arrangement as claimed in claim 6 wherein the units of the third level are formed by mutually interconnected second level units.

8. A network arrangement for interchanging load status information substantially as herein described with reference to the accompanying drawings.

9. A network arrangement as claimed in any one of claims 1 to 8 implementing the maximal flow techniques of 44470/99.

1/3

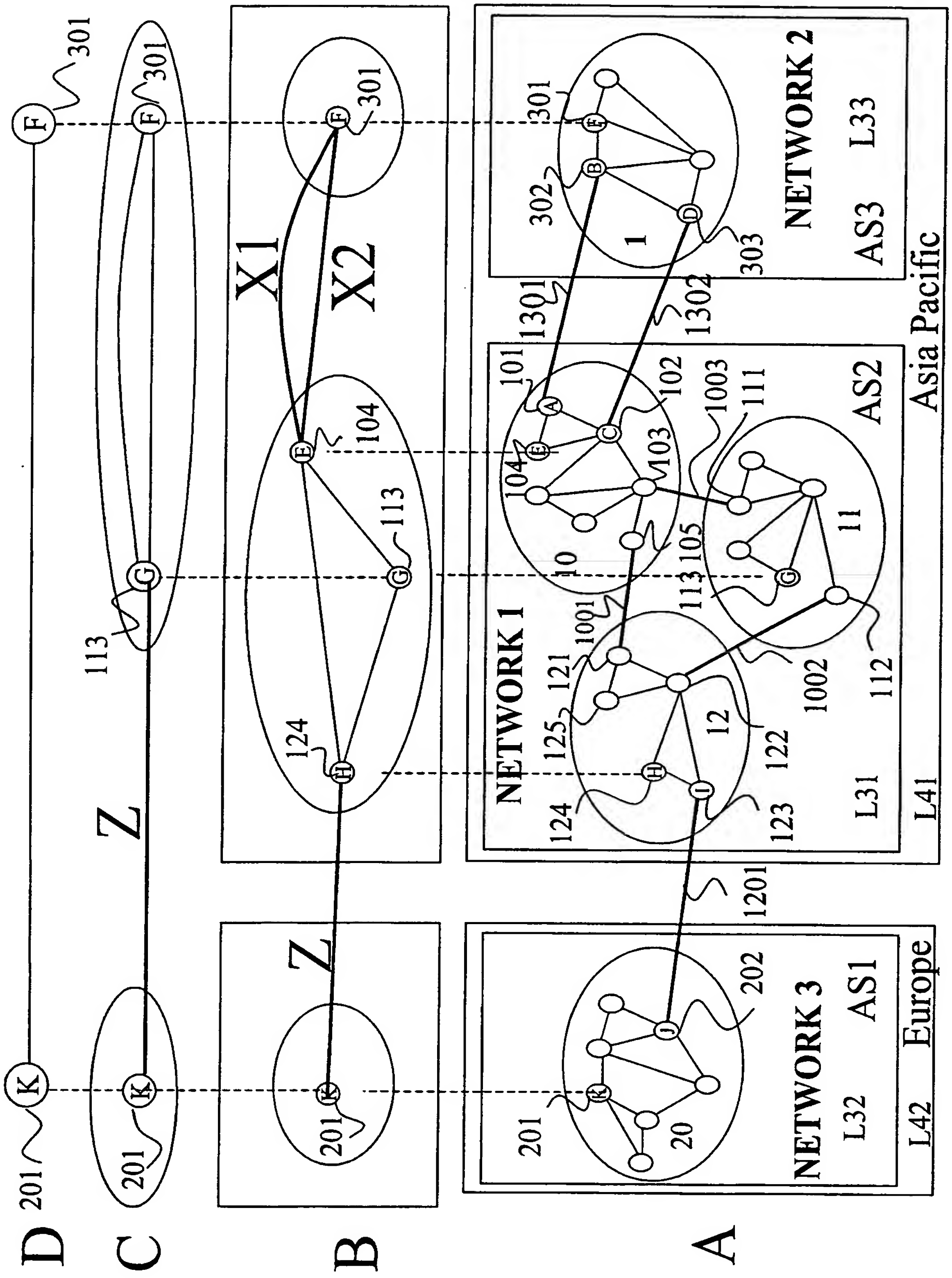

D
C
B
A
Z
X1
X2
NETWORK 1
NETWORK 2
NETWORK 3
AS1
AS2
AS3
L31
L32
L33
L41
L42
Asia Pacific
Europe
201
202
301
302
303
101
102
103
104
105
111
112
113
121
122
123
124
125
1001
1002
1003
1201
1301
1302
10
11
12
20
1
K
F
G
E
H
A
B
C
D
I
J

FIGURE 1

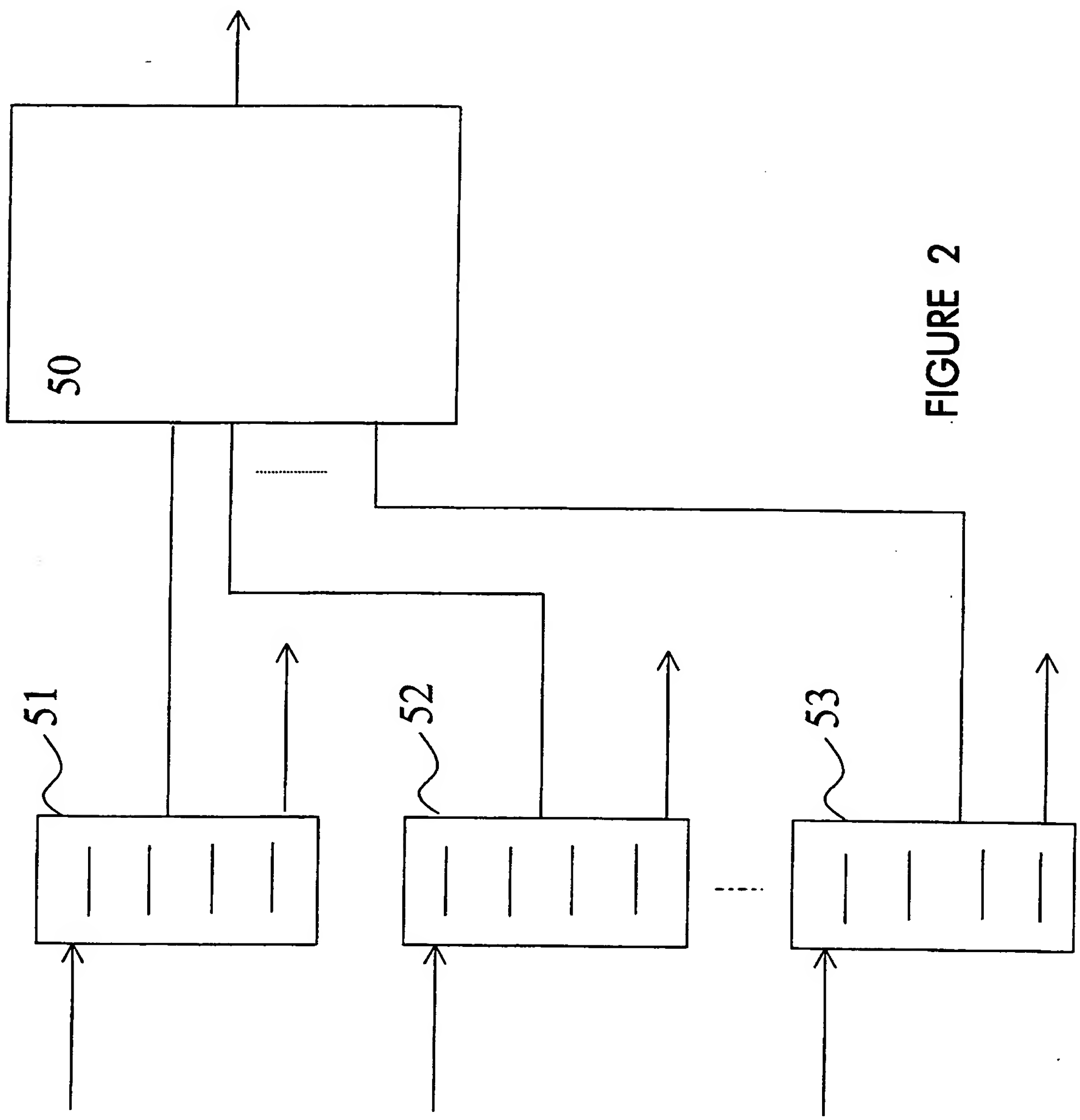


FIGURE 2

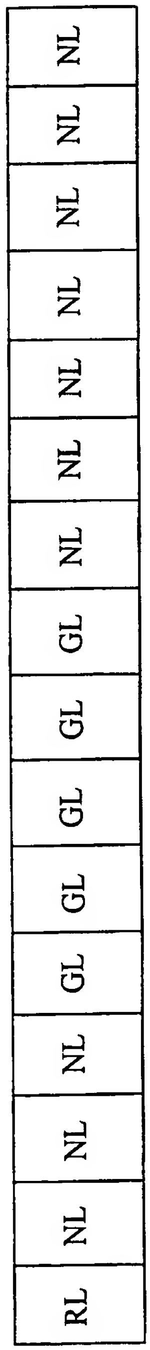


FIGURE 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU 00/01023

A.	**CLASSIFICATION OF SUBJECT MATTER**
Int Cl[7]:	H04L 12/00
According to International Patent Classification (IPC) or to both national classification and IPC	
B.	**FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)
WHOLE IPC

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
AU: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WPAT: (NETWORK + or LAN or Wan) and (HIERARCH + or TREE or LAYER) and (DISTRIBUTION? or LOAD or TRAFFIC or FLOW or CONGESTION) and (STATUS or CONDITION or STATE)
INSPEC: DITTO

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5483533A (KUBA) 9 January 1996 Whole document	1-9
A	US 5537468A (HARTMANN) 16 July 1996 Whole document	1-9
A	US 5872773A (KATZELA et al.) 16 February 1999 Whole document	1-9

[X] Further documents are listed in the continuation of Box C

[X] See patent family annex

* Special categories of cited documents:

- "A" Document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed
- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search 29 September 2000	Date of mailing of the international search report 12 OCT 2000
Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200 WODEN ACT 2606 AUSTRALIA E-mail address: pct@ipaustralia.gov.au Facsimile No.: (02) 6285 3929	Authorized officer **JUZER KHANBHAI** Telephone No.: (02) 6283 2176

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU 00/01023

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	DE 19742582 C1 (SIEMENS AG) 29 April 1999 Whole document	1-9
A	US 5905871A (BUSKENS et al.) 18 May 1999 Whole document	1-9

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/AU 00/01023

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member			
US	5483533	JP	7087112		
US	5537468	EP	608279	WO	9308666
US	5872773	NONE			
DE	19742582	NONE			
US	5905871	NONE			

END OF ANNEX